Overview and explanation of all processing that was done on the dataset.

1. Dropping Columns

A new DataFrame, ‘df1’, is created by dropping the ‘reservation_status_date’ column from the original DataFrame ‘df’.

2. Identifying Categorical Variables

Identifying the names of all categorical columns in ‘df1’ whose data type is the object.

3. Concatenating Lists

Adding columns with numerical representations of categorical variables to the list ‘cate’.

4. Fill NA Values in ‘agent’ and ‘company’

Filling missing values in the ‘agent’ and ‘company’ columns with 0s.

5. Creating New Binary Features

Creating new binary features ‘in_company’ and ‘in_agent’ to classify rows based on whether the ‘company’ and ‘agent’ columns are 0 or not.

6. Creating the ‘same_assignment’ Feature

Creating a new binary feature, ‘same_assignment’, that is 'Yes' if the reserved room type is the same as the assigned room type, and 'No' otherwise.

7. Further Dropping of Columns

Dropping the columns ‘reserved_room_type’, ‘assigned_room_type’, ‘agent’, and ‘company’ from ‘df1’.

8. Re-encoding ‘is_repeated_guest’

Re-encoding the ‘is_repeated_guest’ column with 'YES' for 1 and 'NO' for 0.

9. Filling Missing Values in ‘country’

Filling the missing values in the ‘country’ column with the mode of that column.

10. Encoding Categorical Features

Encoding the categorical features using Ordinal Encoder.

11. Imputing Missing Values in ‘children’

Filling missing values in the ‘children’ column with its mode.

12. Standardizing Continuous Features

Scaling the continuous variables using Standard Scaler.

13. Calculating Correlations

Calculating the correlation of all numerical columns with the ‘is_canceled’ column.

14. Visualization of Correlations

Creating a horizontal bar plot to visualize the absolute correlation values between the ‘is_canceled’ column and other numerical columns.

15. Dropping Additional Column

Creating a new DataFrame, ‘df2’, by dropping the ‘reservation_status’ column from ‘df1’.